Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7840–7843, August 1992
Biochemistry

The proteins of ISGF-3, the interferon α-induced transcriptional activator, define a gene family involved in signal transduction

XIN-YUAN FU*, CHRIS SCHINDLER*, TERESA IMPROTA*, RUEDI AEBERSOLD†, AND JAMES E. DARNELL, JR.*

*Rockefeller University, New York, NY 10021-6399; and †University of British Columbia, Biomedical Research Center and Department of Biochemistry, Vancouver, BC, Canada V6T 1W5

Contributed by James E. Darnell, Jr., May 27, 1992

ABSTRACT **ISGF-3 is a multiprotein transcription factor that is very quickly activated in the cell cytoplasm only after attachment of interferon α to the cell surface. To understand the specific cytoplasmic activation of proteins that move to the nucleus and direct increased transcription of specific genes, we have purified and now report completion of the cloning of cDNA encoding the four proteins of ISGF-3. With all of the sequences available, it is clear that three of these proteins are encoded by members of a previously unrecognized gene family. We suggest that proteins encoded by this gene family serve the function of interpreting the fact that a cell surface receptor has bound its ligand so that specific signal transduction to the nucleus can occur.**

ISGF-3 is a multiprotein complex consisting of a 48-kDa protein capable of weak site-specific DNA binding and three proteins, 113, 91 and 84 kDa, that become activated in the cell cytoplasm by attachment of interferon α (IFN-α) to cells. The ISGF-3 proteins then enter the nucleus and join with a 48-kDa protein to form a tight DNA-binding complex (1–4). Because second-messenger mimics do not activate ISGF-3 and because IFN-α activates the factor within seconds in the cell cytoplasm, it was proposed that a direct protein–protein interaction was involved in the signal transduction that results in specific gene stimulation (4).

We have recently cloned two cDNAs from one gene that encodes the 91- and 84-kDa proteins (5), and antisera against those proteins reacted with the ISGF-3 DNA complex. We have also cloned the cDNA encoding the 48-kDa DNA-binding protein (6). That cDNA sequence shows some similarity to the ISGF-2 (IRF-1) DNA (7, 8) and encodes a protein different from the three larger ISGF-3 proteins. We have now cloned and sequenced a cDNA encoding the 113-kDa protein,‡ and antiserum against this protein also reacts with the ISGF-3 DNA complex. Comparison of the sequences of the 113-kDa and the 91/84-kDa proteins reveals that they are encoded by closely related but distinct genes. Inspection of the 113-kDa and 91/84-kDa sequences reveals heptad leucine repeats in a potentially helical region (9, 10) that could account for complex formation after IFN-α treatment. In addition, one short conserved region of the two genes encodes a sequence similar to the Src homology 2 (SH2) domains noted in various tyrosine kinases and substrates for these enzymes (11). We speculate that the genes encoding the 113- and 91/84-proteins are members of a gene family whose products serve directly to receive information that a specific receptor has bound its ligand and subsequently transduce this information to the nucleus.

MATERIALS AND METHODS

The purification of ISGF-3 for microsequencing (12) was described earlier (1, 5). Sequences obtained by microsequencing of the ISGF-3 113-kDa component are underlined in Fig. 1. Based on peptide E, we designed a degenerate oligonucleotide, AAYACIGARCCIATGGARATYATT, which was used to screen a cDNA library basically as described (13). Briefly, the degenerate oligonucleotides were labeled by polynucleotide kinase with [γ-^{32}P]ATP and hybridizations were carried out overnight at 40°C in 6× SSTE (0.9 M NaCl/60 mM Tris·HCl, pH 7.9/6 mM EDTA)/0.1% SDS/2 mM $Na_4P_2O_7$/6 mM KH_2PO_4/10 × Denhardt's solution (13) in the presence of salmon sperm DNA (100 μg/ml). The nitrocellulose filters then were washed four times (10 min per wash) with the same hybridization conditions but without labeled probe and salmon sperm DNA. Autoradiography was carried out at −80°C with an intensifying screen for 48 h. A PCR product (14) was also obtained by using oligonucleotides designed according to peptides D and E. The sequence of this PCR product was identical to a region in clone f11. The full-length cDNA clones of the 113- and 91-kDa proteins were transcribed *in vitro* and transcribed RNAs were translated *in vitro* with rabbit reticulate lysate (Promega; conditions as described in the Promega protocol).

In vitro transcription and translation of cDNA sequences contained in bacterial plasmids was carried out with a Promega kit according to the manufacturer's instructions.

Antibodies to two segments of the 113-kDa protein (amino acids 500–650 and 323–527) were prepared against bacterial fusion proteins expressed from the pGEX-2T vector (15). One milligram of antigen was used for immunization and two booster injections 2 weeks apart were given. Western blots and precipitations were carried out by standard methods (16).

Northern blot analysis was carried out as described (8, 13).

RESULTS

We purified sufficient ISGF-3α to allow peptide sequence to be obtained and to begin cloning of all four ISGF-3 proteins (1, 5). From SDS gels of highly purified ISGF-3, the 113-kDa band was identified, excised, and subjected to cleavage and peptide sequence analysis. Five peptide sequences (A–E) were obtained (Fig. 1*A*). We designed degenerate oligonucleotide probes that would encode these peptides. The probes were radiolabeled to screen a human cDNA library for clones that might encode the 113-kDa protein. Eighteen positive cDNA clones were recovered from 2.5×10^5 phage plaques with the probe derived from peptide E (Fig. 1*A* and legend). The inserts in two of the clones were completely sequenced. One clone (f11) contained a 3.2-kilobase (kb) cDNA, and a second clone (ka31) a 2.6-kb cDNA segment that overlapped about 2 kb with f11, extended about 2 kb further to the 5′ end, and contained a candidate AUG initiation codon with a well-conserved Kozak sequence (17). In addition to the phage cDNA clones, a PCR product made between oligonucleotides that encoded peptides D and E also yielded a 474-nucleotide fragment that when sequenced was identical with the cDNA

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IFN, interferon; SH2, Src homology 2.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M97934).

A

```
  1:  MAQWEMLQNLDSPFQDQLHQLYSHSLLPVDIRQYLAVWIEDQNWQEAALGSDDSKATMLF
 61:  FHFLDQLNYECGRCSQDPESLLLQHNLRKFCRDIQPFSQDPTQLAEMIFNLLLEEKRILI
121:  QAQRAQLEQGEPVLETPVESQQHEIESRILDLRAMMEKLVKSISQLKDQQDVFCFRYKIQ
                                  Helix 1
181:  AKGKTPSLDPHQTKEQKILQETLNELDKRRKEVLDASKALLGRLTTLIELLLPKLEEWKA
                                  Helix 2
241:  QQQKACIRAPIDHGLEQLETWFTAGAKLLFHLRQLLKELKGLSCLVSYQDDPLTKGVDLR
                                  Helix 3
301:  NAQVTELLQRLLHRAFVVETQPCMPQTPHRPLILKTGSKFTVRTRLLVRLQEGNESLTVE
361:  VSIDRNPPQLQGFRKFNILTSNQKTLTPEKGQSQGLIWDFGYLTLVEQRSGGSGKGSNKG
421:  PLGVTEELHIISFTVKYTYQGLKQELKTDTLPVVIISNMNQLSIAWASVLWFNLLSPNLQ
481:  NQQFFSNPPKAPWSLLGPALSWQFSSYVGRGLNSDQLSMLRNKLFGQNCRTEDPLLSWAD
541:  FTKRESPPGKLPFWTWLDKILELVHDHLKDLWNDGRIMGFVSRSQERRLLKKTMSGTFLL
601:  RFSESSEGGITCSWVEHQDDDKVLIYSVQPYTKEVLQSLPLTEIIRHYQLLTEENIPENP
661:  LRFLYPRIPRDEAFGCYYQEKVNLQERRKYLKHRLIVVSNRQVDELQQPLELKPEPELES
721:  LELELGLVPEPELSLDLEPLLKAGLDLGPELESVLESTLEPVIEPTLCMVSQTVPEPLQG
781:  PVSQPVPEPDLPCDLRHLNTEPMEIFRNCVKIEEIMPNGDPLLAGQNTVDEVYVSRPSHF
841:  YTDGPLMPSDF
```

B

```
113 kDa  MAQWEMLQNLDSPFQDQLHQLYSHSLLPVDIRQYLAVWIEDQNWQEAALGSDDSKATMLF
91/84 kDa MSQWYELQQLDSKFLEQVHQLYDDS-FPMEIRQYLAQWLEKQDWEHAA--NDVSFATIRF

61   FHFLDQINYECGRCSQDPESLLLQHNLRKFCRDIQP-FSQDPTQLAEMIFNLLLEEKRIL
57   HDLLSQIDDQYSRFSLE-NNFLLQHNIRKSKRNLQDNFQEDPIQMSMIIYSCLKEERKIL

120  IQAQRAQLEQGEPVLETPVESQQHEIESRILDLRAMMEKLVKSISQLKDQQDVFCFRYK-
117  ENAQRFNQAQSGNIQSTVMLDKQKELDSKVRNVKDKVMCIEHEIKSLEDLQDEYDFKCKT

179  IQAKGKTPS--LDPHQTKEQKILQETLNELDKRRKEVLDASKALLGRLTTLIE--LLLPK
177  LQNREHETNGVAKSDQKQEQLLLKKMYLMLDNKRKEVVHKIIELL-NVTELTQNALINDE

235  LEEWKAQQQKACIRAPIDHGLEQLETWFTAGAKLLFHLRQLLKELKGLSCLVSYQDDPLT
236  LVEWKRRQQSACIGGPPNACLDQLQ----------QVRQQLKKLEELEQKYTYEHDPIT

295  KGVDLRNAQVTELLQRLLHRAFVVETQPCMPQTPHRPLILKTGSKFTVRTRLLVRLQEGN
285  KNKQVLWDRTFSLFQQLIQSSFVVERQPCMPTHPQRPLVLKTGVQFTVKLRLLVKLQELN

355  ESLTVEVSIDRNPPQ---LQGFRKFNILTSNQKTLTPEKGQSQGLIWDFGYLTLVEQRSG
345  YNLKVKVLFDKDVNERNTVKGFRKFNILGTHTKVMNMEESTNGSLAAEFRHLQLKEQKNA

412  GSGKGSNKGPLGVTEELHIISFTVKYTYQGLKQELKTDTLPVVIISNMNQLSIAWASVLW
405  GT--RTNEGPLIVTEELHSLSFETQLCQPGLVIDLETTSLPVVVISNVSQLPSGWASILW

472  FNLLSPNLQNQQFFSNPRKAPWSLLGPALSWQFSSYVGRGLNSDQLSMLRNKLFGQNCRT
463  YNMLVAEPRNLSFFLTPPCARWAQLSEVLSWQFSSVTKRGLNVDQLNMLGEKLLGPNASP

532  EDPLLSWADFTKRESPPGKLPFWTWLDKILELVHDHLKDLWNDGRIMGFVSRSQERRLLK
523  DG-LIPWTRFCKENINDKNFPFWLWIESILELIKKHLLPLWNDGCIMGFISKERERALLK

592  KTMSGTFLLRFSESS-EGGITCSWVEH-QDDDKVLIYSVQPYTKEVLQSLPLTEIIRHYQ
582  DQQPGTFLLRFSESSREGAITFTWVERSQNGGEPDFHAVEPYTKKELSAVTFPDIIRNYK

650  LLTEENIPENPLRFLYPRIPRDEAFGCYY------QEKVNLQERR--KYLKHRLIVVSNR
642  VMAAENIPENPLKYLYPNIDKDHAFGKYYSRPKEAPEPMELDGPKGTGYIKTELISVSEV

702  QVDELQQPLELKP
702  HPSRLQTTDNLLP
```

FIG. 1. (*A*) Full-length amino acid sequence of 113-kDa protein component of ISGF-3α. Polypeptide sequences (A–E) derived from protein microsequencing of purified 113-kDa protein are underlined. The full-length 113-kDa protein contains 851 amino acids. Three major helixes in the amino-terminal region were predicted by the methods of both Chou and Fasman (11) and Garnier *et al.* (12) and are shown in boxes. At the carboxyl-terminal end, a highly negatively charged domain is present. All negative charged residues are blackened and positively charged residues are shaded. (*B*) Comparison of amino acid sequences of 113-kDa and 91/84-kDa protein shows 42% identity in amino acid residues in the overlapping 715-amino acid sequence shown. When a heligram structure was drawn this helix was found to be amphipathic (data not shown). Several tyrosine residues (boxed) are conserved in both proteins. The conserved region at residues 596–606 (numbering according to 113-kDa sequence) is similar to regions in SH2 domains (9).

clone in this region. A combination of the clones f11 and ka31 revealed an open reading frame capable of encoding a polypeptide of 851 amino acids (Fig. 1*A*). These two clones were joined within their overlapping region, and RNA transcribed from this recombinant clone was translated *in vitro* to yield a polypeptide that migrated with a nominal molecular mass of 105 kDa in an SDS/polyacrylamide gel (Fig. 2*A*). An appropriate clone encoding the 91-kDa protein (5) was also transcribed and the RNA was translated in the same experiment. Since both the apparently complete cDNA clones for the 113-kDa protein and the 91-kDa protein produce RNAs that when translated into proteins migrate somewhat faster than the proteins purified as ISGF-3 components, it is possible that the proteins undergo posttranslational modification in the cell that causes them to be slightly retarded during electrophoresis. When a 660-base-pair cDNA probe for the mRNA encoding the 113-kDa protein was used in a Northern analysis, a single 4.8-kb mRNA was observed (Fig. 2*B*).

We have no independent assay for the activity of the 113-kDa protein (or indeed for any of the cytoplasmically activated ISGF-3 proteins), but we do know that the protein is part of a DNA-binding complex that can be detected by an electrophoretic mobility shift assay (1). Antibodies to DNA-binding proteins are known to affect the formation or migration of such complexes. Therefore, antiserum to a polypeptide segment (amino acid residues 323–527) fused with bacterial glutathione-*S*-transferase (8) was raised in rabbits to determine the reactivity of any proteins within ISGF-3 with the antibody. A Western blot analysis showed that the antiserum reacted predominantly with a 113-kDa protein both in the ISGF-3 fraction purified by specific DNA affinity chromatography (Fig. 3*A*, lane 1) and in crude cell extract (lane 2). The weak reactivity to lower protein bands was possibly due to degradation of the 113-kDa protein. Most importantly, the antiserum specifically removed almost all of the characteristic ISGF-3 gel-shift complex, leaving some of the oligonucleotide probe in "shifted-shift" complexes. All of the DNA complexes were specifically removed by competition with a 50-fold molar excess of the oligonucleotide binding site [the interferon stimulation response element, ISRE (1–3)] for ISGF-3 (Fig. 3*B*). Notably, this antiserum had no effect on the faster migrating shift band produced by the 48-kDa (labeled ISGF-3γ) component alone (Fig. 3*B*). This faster moving gel-shift band is labeled ISGF-3γ because this protein is increased in HeLa cells by pretreatment with IFN-γ (1, 2, 19). Thus it appeared that the antiserum to the 113-kDa fusion product did indeed react with a specific protein that was part of the complete ISGF-3 complex. Similar evidence with other antibodies has also been obtained that the 91- and 84-kDa proteins were correctly identified as part of ISGF-3 and show no crossreactivity with the 48-kDa protein (5).

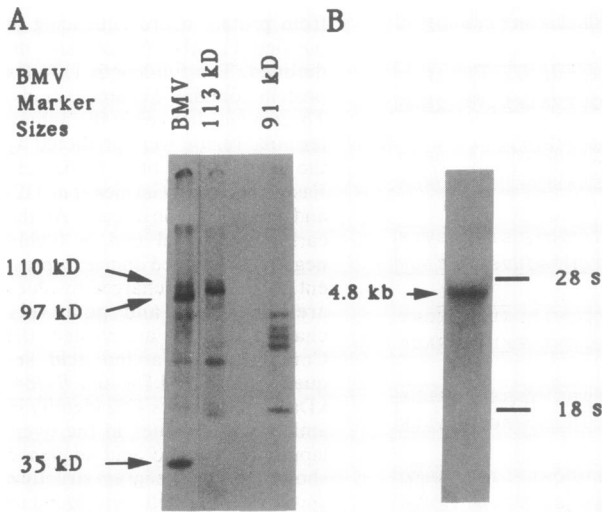

FIG. 2. (*A*) *In vitro* transcription/translation products of cDNAs encoding the 113-kDa and 91-kDa proteins. The mRNA of brome mosaic virus (BMV) was simultaneously translated to give protein size markers. (*B*) Northern blot of cytoplasmic mRNAs isolated from HeLa cells pretreated overnight with IFN-γ. Positions of 28S and 18S rRNAs provided size markers.

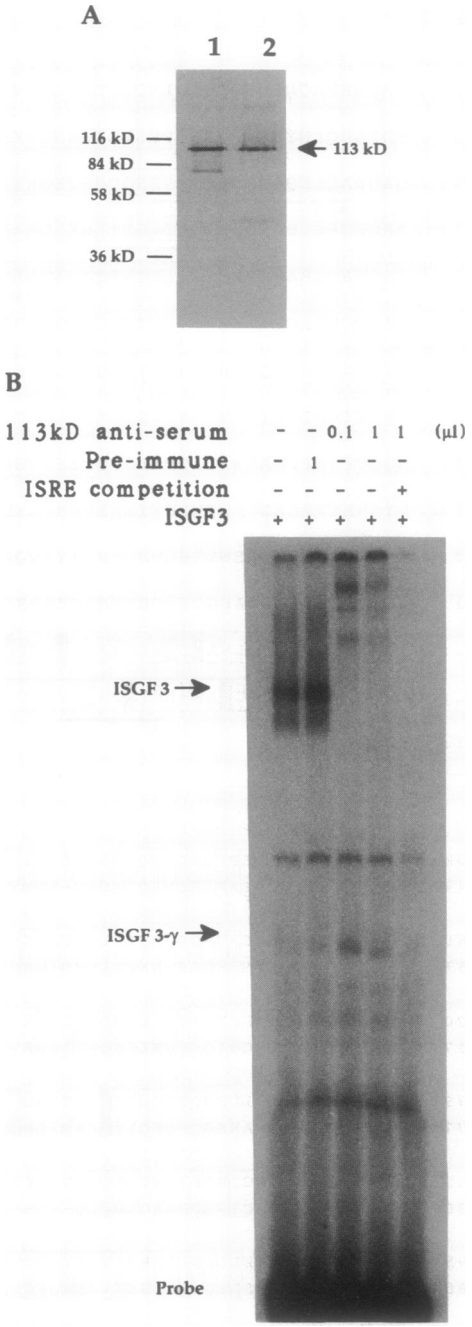

FIG. 3. (*A*) An antiserum raised against a polypeptide segment (18), amino acids 500–650, of the 113-kDa protein recognizes by Western blot analysis a 113-kDa protein both in crude extracts and in purified ISGF-3 (16). (*B*) The anti-113-kDa antiserum affects the ISGF-3 shift complex. Preimmune serum or the 113-kDa antiserum was added to the gel-shift reaction mixture (1) at room temperature for 20 min with partially purified ISGF-3 protein. A 50-fold excess of unlabeled insulin stimulation response element (ISRE) probe was included in one reaction mixture to show specificity of the gel-shift complexes.

A detailed comparison between the 113-kDa and 91-kDa sequences followed (Fig. 1*B*). While the nucleotide sequence showed only a distant relationship between the two proteins, there were long stretches of amino acid identity. These conserved regions were scattered throughout almost the entire 715-amino acid length encoded by the 91/84 clone. It was particularly striking that the regions of the 113-kDa sequence corresponding to amino acids 1–48, 317–353, and 654–678 were 60–70% identical to corresponding regions of the 91-kDa sequence. Thus, the genes encoding the 113-kDa and 91/84-kDa proteins are closely related but not identical.

Through examination for possible consensus sequences that might reveal subdomain structures in the 113-kDa or 91/84-kDa sequence found that both proteins contained regions of heptad leucine repeats (20, 21) that would have the ability to form a coiled-coil structure. These were located between amino acids 210 and 245 in the 113-kDa protein and between 209 and 237 in the 91/84-kDa protein. In both the 113-kDa and the 91/84-kDa sequences, four out of five possible heptad repeats were leucine and one was valine. Domains of this type might provide a protein surface that encourages homo- or heterotypic protein interactions like those which have been observed in several other transcription factors (20, 21). An extended acidic domain was located at the carboxyl terminus of the 113-kDa protein but not in the 91-kDa protein (Fig. 1*A*), possibly implicating the 113-kDa protein in gene activation (18, 22). Finally, a sequence resembling a region within SH2 domains (9) was noted between residues 596 and 606 (in the 113-kDa sequence). This is exactly conserved in the 91/84-kDa sequence and is preceded in the sequence by a tryptophan, an arginine, and two leucines within the 20 upstream amino acids. The relevance of these sequences will require deletion or mutation in future experiments.

Comparison at moderate or high stringency with the available data bases (December 1991) showed no long sequences like the 113-kDa or the 91/84-kDa sequence. Preliminary genomic screening experiments, however, have indicated that there may be other family members with different sequences recoverable from a human genomic DNA library (S. Qureshi and J.E.D., unpublished work). Thus, it appears that the 113- and 91/84-kDa sequences may represent the first two members to be cloned of a larger family of proteins.

We hypothesize that the 113- and 91/84-kDa proteins act in the cytoplasm as signal transducers, somehow interacting with the internal domain of a liganded IFN-α receptor or some protein associated with the receptor. Such an arrangement would conserve the high degree of specificity inherent in ligand–receptor interactions and in the individually specific transcriptional responses to specific ligands (4). Many experiments lie ahead before this general hypothesis can be crucially tested. Recent experiments have indicated that inhibitors of protein kinases can prevent ISGF-3 complex formation (23, 24). However, neither the IFN-α nor the IFN-γ receptors that have so far been cloned have intrinsic kinase activity (25, 26). Perhaps either a second receptor chain with kinase activity or a separate kinase bound to a liganded receptor could be a part of a complex that would convey signals to the waiting ISGF-3 proteins at the inner surface of the plasma membrane.

The cloning of the cDNAs for the ISGF-3 proteins and production of targeted antisera to the proteins should allow us to explore the domains involved in protein–protein interactions, as well as the possible phosphorylation and membrane association, even if transient, of the 113- and 91/84-kDa proteins.

We wish to thank Anthony Canova and Edward Bures for excellent technical assistance, Weimin Zhong for help with sequence analysis, and Dr. Uli Nielsch for help with Northern blotting. This work was supported by a grant from the American Cancer Society (MV424U) and the National Cancer Institute of Canada. X.-Y.F. was a Fellow of the Cancer Research Institute/Fan Fox and Leslie R. Samuels Foundation (New York).

1. Fu, X., Kessler, D. S., Veals, S. A., Levy, D. E. & Darnell, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8555–8559.
2. Levy, D. E., Kessler, D. S., Pine, R., Reich, N. & Darnell, J. E. (1989) *Genes Dev.* **3**, 1362–1371.
3. Kessler, D. S., Veals, S. A., Fu, X.-Y. & Levy, D. E. (1990) *Genes Dev.* **4**, 1753–1765.
4. Levy, D. & Darnell, J. E. (1991) *New Biol.* **2**, 923–928.
5. Schindler, C., Fu, X.-Y., Improta, T., Aebersold, R. & Darnell, J. E., Jr. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7836–7839.
6. Veals, S. A., Schindler, C., Leonard, D., Fu, X.-Y., Aebersold, R., Darnell, J. E. & Levy, D. (1992) *Mol. Cell. Biol.*, in press.
7. Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T. & Taniguchi, T. (1988) *Cell* **54**, 903–913.
8. Pine, R., Decker, T., Kessler, D., Levy, D. & Darnell, J. E., Jr. (1990) *Mol. Cell. Biol.* **10**, 2448–2457.
9. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251–276.
10. Garnier, D., Oglethorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
11. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. (1991) *Science* **252**, 668–674.
12. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6970–6974.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Norman, C. (1988) *Cell* **55**, 989–1003.
15. Smith, D. B., Davern, K. M., Board, P. G., Tiu, W. U., Garcia, E. G. & Mitchell, G. F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8703–8707.
16. Harlow, E. & Lane, D. (1988) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
17. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
18. Hope, I. A. & Struhl, K. (1986) *Cell* **46**, 885–894.
19. Levy, D. E., Lew, D. E., Decker, T., Kessler, D. S. & Darnell, J. E. (1990) *EMBO J.* **9**, 1105–1111.
20. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
21. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246**, 911–916.
22. Ma, J. & Ptashne, M. (1987) *Cell* **48**, 847–853.
23. Reich, N. C. & Pfeffer, L. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8761–8765.
24. Kessler, D. S. & Levy, D. E. (1991) *J. Biol. Chem.* **266**, 23471–23476.
25. Uze, G., Lutfalla, G. & Gresser, I. (1990) *Cell* **60**, 225–234.
26. Aguet, J. M., Dembic, Z. & Merlin, G. (1988) *Cell* **55**, 273–280.